Final Technical Report for Hybrid Algorithms and Oblivious Decision Graphs using $\mathcal{MLC}$++

Ronny Kohavi (Nils J. Nilsson, Principal Investigator)
Stanford University

February 29, 1996

This report details the research and development work done on $\mathcal{MLC}$++ under ONR grant N00014-95-1-0669.

1 Overview of $\mathcal{MLC}$++

$\mathcal{MLC}$++ is a Machine Learning library of C++ classes. General information about the library, including source code, can be obtained through the World Wide Web at URL

`http://robotics.stanford.edu:/users/ronnyk/mlc.html` .

Over 350 different sites have copied the $\mathcal{MLC}$++ kit, and machine learning research in the robotics lab at Stanford is enhanced through the use of the library.

2 Summary of Results

As detailed in the statement of work for the grant, three main projects were proposed:

1. Hybrid decision tree and nearest-neighbor.
2. Stacking and Bagging.
3. Oblivious decision graphs.

We now describe the specific work done and the results obtained.

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
SEATTLE REGIONAL OFFICE
1107 NE 45TH STREET, SUITE 350
SEATTLE WA 98105-4631

IN REPLY REFER TO:

4330
ONR 247
11 Jul 97

From: Director, Office of Naval Research, Seattle Regional Office, 1107 NE 45th St., Suite 350, Seattle, WA 98105

To: Defense Technical Center, Attn: P. Mawby, 8725 John J. Kingman Rd., Suite 0944, Ft. Belvoir, VA 22060-6218

Subj: RETURNED GRANTEE/CONTRACTOR TECHNICAL REPORTS

1. This confirms our conversations of 27 Feb 97 and 11 Jul 97. Enclosed are a number of technical reports which were returned to our agency for lack of clear distribution availability statement. This confirms that all reports are unclassified and are "APPROVED FOR PUBLIC RELEASE" with no restrictions.

2. Please contact me if you require additional information. My e-mail is *silverr@onr.navy.mil* and my phone is (206) 625-3196.

ROBERT J. SILVERMAN

STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-2140

DEPARTMENT OF COMPUTER SCIENCE

Telephone:
(415) 723-2273

March 4, 1996

Program Manager/Officer ONR: 311
Michael O. Shneier
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

Re.: N00014-95-1-0669

Dear Mr. Shneier:

Enclosed please find three copies of the final technical report for the above grant.

Sincerely,

Jutta E. McCormick

Jutta E. McCormick
Administrator, Robotics Laboratory

cc: Administrative Grants Officer, ONR
Seattle, WA (1 copy)
Director, Naval Research Laboratory;
Washington, D.C. (1 copy)
Defense Technical Information Center,
Alexandria, VA (2 copies)
Sponsored Projects Office, Stanford

2.1 Hybrid Decision Tree and Nearest-Neighbor

The proposal called for implementation of a hybrid approach that uses decision trees for the nominal features and uses nearest-neighbor algorithms at the leaves.

The algorithm was implemented in $\mathcal{MLC}$++, and its accuracy outperformed both decision tree algorithms and nearest-neighbor approaches on some artificial datasets that originally motivated this idea. However, performance on real datasets from the UC Irvine repository (Murphy & Aha 1995) did not improve over these approaches in many cases and degraded in some.

Our analysis revealed that while the decision tree algorithm provides a decomposition of the problem, the main problem is that the training set is fragmented because of the decision tree splits and the space for the nearest-neighbor algorithm becomes sparse. The tradeoff between the extra representation power and fewer instances per node does not seem useful in many datasets.

We are currently experimenting with other hybrid approaches that involve parametric algorithms, such as Naive-Bayes (Langley, Iba & Thompson 1992, Duda & Hart 1973), which (because of their limited representation power) do not suffer from the curse of dimensionality in high dimensional spaces, and might therefore be more suitable for this approach.

2.2 Stacking and Bagging

Stacking (Wolpert 1992), sometimes called Bagging (Breiman 1994), averaging (Perrone 1993) or ensembles (Krogh & Vedelsby 1995), utilize multiple classifiers that "vote" on the predicted class.

The voting algorithm has been implemented in $\mathcal{MLC}$++, thus providing the ability to wrap around and aggregate any existing algorithm or new algorithm that is implemented in $\mathcal{MLC}$++.

The algorithm was used in demonstrating the bias and variance for zero-one loss functions (Kohavi & Wolpert 1996). Approaches such as nearest-neighbor, which have a bias problem in high-dimensions will not improve. However, improvements can be seen with methods that suffer from high variance, such as decision tree algorithms, or when nonconvergent methods are used (Finnoff, Hergert & Zimmermann 1993).

2.3 Oblivious Decision Graph

Oblivious decision graphs provide a hypothesis space that is easy to understand, yet does not suffer from some of the shortcomings of decision trees. The original work (Kohavi 1994) that

was supported by a previous ONR grant was extended to deal with continuous attributes through discretization.

The results were published in Kohavi's dissertation with graphs depicting some target concepts from the UCI database (Kohavi 1995, Chapter 6). The resulting graphs are much easier to comprehend in most cases than the equivalent decision trees.

3 Summary

We have implemented the three proposed algorithms: hybrid approach, bagging, and oblivious decision graphs.

The hybrid approach led to some negative results and we are considering alternative approaches based on the same idea. The implementation of bagging now provides any algorithm implemented in $\mathcal{MLC}$++ with a wrapper that might improve its performance, especially if the algorithm suffers from instability, as do decision trees. The implementation of oblivious decision graphs was improved with good results that were shown in Kohavi's dissertation.

Copies of papers based on this research are available on the web (as noted in the References), and hard copies can be requested from the author.

References

Breiman, L. (1994), Bagging predictors, Technical Report Statistics Department, University of California at Berkeley.

Duda, R. & Hart, P. (1973), *Pattern Classification and Scene Analysis*, Wiley.

Finnoff, W., Hergert, F. & Zimmermann, H. G. (1993), Improving model selection by nonconvergent methods, *in* "Neural Networks", Vol. 6, pp. 771–783.

Kohavi, R. (1994), Bottom-up induction of oblivious, read-once decision graphs : strengths and limitations, *in* "Twelfth National Conference on Artificial Intelligence", pp. 613–618.

Kohavi, R. (1995), Wrappers for Performance Enhancement and Oblivious Decision Graphs, PhD thesis, Stanford University, Computer Science department. Available off: http://robotics.stanford.edu/users/ronnyk. STAN-CS-TR-95-1560.

Kohavi, R. & Wolpert, D. H. (1996), Bias plus variance decomposition for zero-one loss functions, Submitted. Available off http://robotics.stanford.edu/users/ronnyk.

Krogh, A. & Vedelsby, J. (1995), Neural network ensembles, cross validation, and active learning, *in* "Advances in Neural Information Processing Systems", Vol. 7, MIT Press.

Langley, P., Iba, W. & Thompson, K. (1992), An analysis of bayesian classifiers, *in* "Proceedings of the tenth national conference on artificial intelligence", AAAI Press and MIT Press, pp. 223–228.

Murphy, P. M. & Aha, D. W. (1995), UCI repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html.

Perrone, M. (1993), Improving regression estimation: averaging methods for variance reduction with extensions to general convex measure optimization, PhD thesis, Brown University, Physics Dept.

Wolpert, D. H. (1992), "Stacked generalization", *Neural Networks* **5**, 241–259.